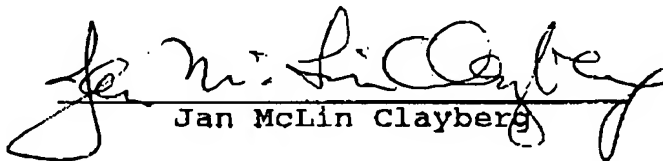# BEST AVAILABLE COPY

May 22, 2006

<u>DECLARATION</u>

The undersigned, Jan McLin Clayberg, having an office at 5316 Little Falls Road, Arlington, VA 22207-1522, hereby states that she is well acquainted with both the English and German languages and that the attached is a true translation to the best of her knowledge and ability of the specification and claims of United States Provisional Patent Application Serial No. 60/465,692, filed in German, in the name of OBENDORF, M., et al.

The undersigned further declares that the above statement is true; and further, that this statement was made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or document or any patent resulting therefrom.

Jan McLin Clayberg

METHOD FOR DETERMINING THE HORMONAL EFFECT OF SUBSTANCES

The invention relates to a method for determining the hormonal effect of substances and to a method for determining disturbances in the co-modulation mechanism of nuclear receptors (NR). The invention also relates to the use of EWS protein (EWS) or of EWS derivatives as well of nucleic acids which code for them.

In the assessment of substances for possible pharmaceutical use, it is generally usual to check these substances for any possible hormonal action, especially for an androgenic or antiandrogenic activity that may be present. When pharmacologically active substances are administered, knowledge about their hormonal effects, especially androgenic or antiandrogenic effects, are important for the sake of assessing any side effects that occur. For testing the hormonal effect of substances, methods are for instance used in which the capability of the substances to bind to the hormone receptors and activate their transcription activity is measured.

However, knowledge about the hormonal effects of substances is of interest not only in potential pharmaceuticals but also in nonpharmaceutical substances, since for many substances present in the environment it is assumed that in some portions of the population they may have an androgenic or antiandrogenic or estrogenic or antiestrogenic activity. Possibly, this may cause an unwanted, harmful effect.

A particular difficulty resides in identifying and characterizing effects on the actions mediated by steroid hormones, since the signal cascades and signal networks that in the final analysis control the hormone-mediated transcription regulation are especially complex in their embodiment here. The reason for this is the very similar embodiment of the DNA target sequences to which the various steroid hormone receptors bind after ligand activation. As a result, for triggering a targeted response, the nuclear receptors are dependent on the interaction with special co- factors which among other effects amplify the specificity of the receptor-mediated transcription activation.

For identifying substances which act on certain hormone-induced signal paths, test systems and test methods are therefore necessary that are capable of purposefully detecting the function of individual components of the cellular signal network for mediating

There is accordingly a need for a method with which a statement about the hormonal effect of substances can be made in a reliable, sensitive, simple, inexpensive way, and quickly. The previously known methods do not meet this need.

It is therefore the object of the present invention to furnish a method with which a statement about the hormonal effect of the substances to be tested can be made in a reliable, sensitive, simple, inexpensive way, and quickly.

This object is attained according to the invention by a method for determining the hormonal effect of substances, having the steps of putting EWS protein (EWS) or a derivative thereof in contact with an NR (nuclear receptor) or a derivative thereof and a test substances, and

a) determining the influence of the test substance on the binding of EWS protein or its derivative and NR or its derivative; or

b) determining the influence of the test substance on the ligand-induced activity of the NR.

The term derivative of a protein or polypeptide is understood in the context of the present invention to mean any variant of the protein or polypeptide obtained by amino acid deletion, substitution, insertion, inversion, addition, or exchange. Those derivatives (functional derivatives) that have maintained the capacity of the unaltered protein/polypeptide to influence the activity of other proteins, for instance, or at least to bind them are preferred.

The invention is based on the surprising finding that the EWS protein and

derivatives derived from it (hereinafter called EWS) have the capability of interacting with nuclear receptors (or derivatives thereof) and modulating their activity.

The superfamily of nuclear receptors (NRs), which includes more than fifty different proteins, is a group of related transcription factors, which control the transcription of the particular target gene as a reaction to specific ligands, such as hormones. The family can be subdivided into a plurality of subfamilies in accordance with certain characteristics, such as the dimerization status, type of ligand, or structure of the DNA reaction element (Beato et al, 2000, Human Reproduct. Update, 6, 225-236). A characteristic feature of NRs is the matching structure of the functional domain (with the designations A-F) with a highly variable, only poorly preserved N-terminal region with an autonomous constitutive activation function (AF-1), a strongly preserved DNA binding domain (DBD), which is responsible for the detection of special DNA reaction elements, and two zinc finger motifs, a variable hinge domain, and a preserved multifunctional C-terminal ligand binding domain (LBD) with a dimerization- and ligand- dependent transactivation function (AF-2). This is followed by the farthest C-terminally located region, whose function is unknown, and which is missing in such receptors as PR (progesterone receptor), PPAR (peroxisome proliferator- activated receptor) and RXR (retinoid-X receptor) (Mangelsdorf & Evans, 1995; Cell, 83, 841-850; Robyr et al, 2000, Mol. Endocrinol., 14, 329-347). For some NRs (such as the androgen receptor (AR)), it was found that the N-terminal region is capable of interacting with the C-terminal region (Brinkmann et al, 1999, J. Steroid Biochem. and Mol. Biol., 69, 307-313). Steroid hormone receptors, such as estrogen receptors (ER), progesterone receptors (PR), glucocorticoid receptors (GR), mineral corticoid receptors (MR), and androgen receptors (AR) bind steroidal ligands, which are derived from pregnenolone, such as the progestins, estrogens, glucocorticoids, and the mineral corticoids, as well as androgens. The ligand binding activates the receptor and controls the expression of corresponding target genes.

EWS is known as a proto-oncogen of Ewing's sarcoma and other neoplasias, such as Sehnen's clear cell sarcoma and aponeuroses, small- and round-cell desmoplastic intraabdominal tumors, and extraskeletal chondrosarcoma (Delattre,

O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., Aurias, A., and Thomas, G., 1992, Nature 359: 162-165, Zucman, J., Delattre, O., Desmaze, C., Epstein, A.L., Stenman, G., Speleman, F., Fletchers, C.D., Aurias, A., and Thomas, G.; 1993, Nature Genet. 4: 341-345, Gerald, W.L., Rosai, J. and Ladanyi, M., 1995, Proc. Natl. Acad. Sci. USA 92: 1028-1032, Laballe, Y., Zucman, J., Stenman, G., Kindblom, L.G., Knight, J., Turc-Carel, C., Dockhorn- Dworniszak, B., Mandahl, N., Desmaze, C., Peter, M., Aurias, A., Delattre, O., and Thomas, G., 1995, Hum. Mol. Genet. 4: 2219-2226). In all these tumors, the EWS gene site is rearranged, so that the amino acid end (N-terminus) of the protein fuses to a DNA binding domain of FLI1, ERG1, ATF1, or WT1. This N-terminal end of the fused protein contains the EWS exons 1-7 or 1-8 or 1-9. If the cleavage point is between exon 7 and exon 8, then the EWS proportion of the protein created by the fusion has no matches with the androgen receptor binding domain defined here. Conversely, if the cleavage point is located between exons 8 and 9 or exons 9 and 10, then only 5 or 20 amino acids, respectively, of the two oncogenic EWS fusion proteins match the EWS proportion that contains the androgen receptor binding domain. It can be concluded from this that the rearranged EWS fusion proteins have lost the capability of binding to androgen receptors.

In the analysis of thymus RNA by means of RT-PCR, an EWS variant (EWS1-c) was found, in which 17 amino acids are missing (Fig. 3). Evidently, this is a splice variant, since all the necessary consensus sequences were present at the seams between introns and exons. The result was a shortening of exon 15 (exon 15b). In the prior art, still other splice variants are also known. One of them (Ohno, T., Ouchida, M., Lee, L., Gatalica, Z., Rao, V.N., and Reddy, E.S., 1994, Oncogene 9: 3087-3097) represents an EWS transcript (EWS1-b) that is shortened by up to 200 bp and was found in resting lymphocytes, or lymphocytes stimulated by phytohemagglutinin (PHA). The exons 8 and 9 are missing in the EWS1-b. Another variant (Melot, T., Dauphinot, L., Sevenet, N., Radvanyi, F., Delattre, O. (2001), Eur. J. Biochem. 268, 3482-3489) contains one additional exon 4' between the exons 4 and 5 and is called a brain-specific isoform.

EWS belongs to a group of RNA-binding proteins, to which an implication in

the process of RNA synthesis or processing has been ascribed in the prior art; however, until now, only little was known about the physiological function of the somatic wild type of EWS. In particular, it was not known in the prior art that EWS has the capability of binding nuclear receptors and modulating their activity, as a result of which it must be assigned to the class of NR co- modulators.

An E. coli strain known as Escherichia coli EWS-10 CMX was filed under the number DSM 154417 on January 24, 2003 with the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) [German Collection of Microorganisms and Cell Cultures GmbH]. Escherichia coli EWS-10 CMX contains the full-length EWS-cDNA that can be used in the method of the invention.

So-called co-modulators are a class of proteins which in the activation (co-activators) or repression (co- repressors) of gene transcription serve as bridge molecules between the transcription initiation complex and the NRs (McKenna et al, 1999, Endocr. Rev., 20, 321-347). A co- activator must be capable of amplifying the receptor function and, in the presence of an agonist, of interacting directly with the activation domain of NRs. It must also interact with the basal transcription apparatus, and finally, it must not itself reinforce the basal transcription activity. Most co- modulators, with the aid of one or more LXXLL motifs (NR boxes), interact with the AF-2 domain of NRs, but some co- modulators have also been described which interact with other NR regions (Ding et al, 1998, Mol. Endocrinol. 12, 302-313). Many co-modulators have also been identified that interact in a similar way with a plurality of different NRs.

The method of the invention can be performed either in vitro (that is, for instance as a purely biochemical or biophysical assay, in solution or in suitable solid matrixes, etc.), or partly or entirely in the cellular system. Such different test systems are well known to one skilled in the art.

Preferably, at least one of the method steps is performed in the cellular system, since the effects of the steroid-mediated transcription activation can be represented especially well in the physiological context of the cell. Accordingly, eukaryotic cells are especially suitable for the invention; both primary cells and

established cell lines can be used. The use of established cell lines allows especially good reproducibility and economy; conversely, the use of primary cells largely avoids cell culture artefacts that result from mutation and clonal selection. Prostate cells, nerve cells, glial cells, fibroblasts, blood cells, osteoblasts, osteoclasts, hepatocytes, epithelial cells, or muscle cells are especially suitable.

The hormonal effect determined here (that is, identified, quantified, or characterized) can have the nature of both an activator and an inhibitor, and besides its action on the receptor-co-modulator binding, it can also refer to any other step in the NR action, such as the ligand-induced transactivation, but also locating the nucleus of the NR.

In a preferred embodiment, the method of the invention includes the following steps:

a) First, cells which express EWS or a derivative thereof and NR or a derivative thereof are exposed to the substance to be tested.

b) This is followed by determining the influence of a substance on the interaction between the receptor or its derivative and EWS or its derivative, by measuring the protein-protein interaction or the protein-protein DNA interaction.

The expression of one or both of the components (EWS/derivative on the one hand and NR/derivative on the other) interacting with one another can take place intrinsically in the cell or as a function of transient or stable transfection with suitable expression vectors. Selecting suitable cell types and optionally vector systems is a provision that is familiar to one skilled in the art who is tasked with doing so. For expression in the eukaryotic system, pCMX or pSG5 are examples of suitable vectors.

Measuring the protein-protein interaction between the receptor or derivative and EWS or derivative, or the protein- protein DNA interaction of these components with the DNA target sequence, is done by provisions familiar to one skilled in the art. Suitable techniques for this are for instance the two-hybrid

system, co-immune precipitation, GST pull-down assays, FRET analyses, and ABCD assays or gel retardation assays for analyzing protein-protein DNA interactions.

A preferred embodiment of the method of the invention includes the following steps:

a) cells which express EWS protein or a derivative thereof and NR or a derivative thereof and are transfected with a reporter gene construct are exposed to the ligand of the nuclear receptor and to the substance to be tested;

b) determining the transcription activity of the NR by measuring the reporter gene activity; and

c) comparison with the transcription activity in performing steps a) and b) in the absence of the substance to be tested.

Reporter genes are genes or gene fragments that code for gene products that can be demonstrated as simply as possible, for instance photometrically by color reactions. Reporter genes that are often used are the gene for β- galactosidase, the gene for the alkaline phosphatase, the gene for chloramphenicol-acetyltransferase, the gene for catechol-dioxygenase, the gene for the "green" or "blue fluorescent protein", and various luciferase genes that can cause the cells to emit light. By putting a suitable control element, such as a promoter-enhancer sequence, which is under the control of a certain transcription factor or a certain signal transduction cascade, in front, the activity of the transcription factor or cascade can for instance be determined from the amount of gene product expressed.

Conventionally, such reporter genes are introduced into the cells in suitable vectors, putting the promoter-enhancer sequence of interest in front. To analyze the steroidal activity of substances, all the known NR target sequences are suitable - depending on the NR to be analyzed. One example of such a vector is the MMTV-luciferase vector, which is used to measure the androgenic action of

substances.

Substances with a hormonal effect, preferably an androgenic/antiandrogenic effect, can then be detected from the increased or reduced expression of the reporter gene compared to test batches, without adding the substance to be tested.

For use in the context of the present invention, besides wild type EWS, several examples include EWS derivatives and especially functional EWS derivatives, which have maintained the capability of modulating the activity of at least one nuclear receptor, especially the activity of androgen receptors, or at least to bind them (in a way that can be demonstrated by suitable methods and is not negligible - for instance in protein-protein interaction assays such as EMSA; one skilled in this art can make a differentiation here). The same is analogously true for the NR derivatives; once again, the preferred derivatives are those that have maintained the capability of being modulated or at least bound by EWS or its functional derivatives.

EWS and EWS-coding nucleic acids are already known in the prior art (see above). For use in the context of the present invention, an EWS coded by the nucleic acid in accordance with sequence ID No. 1, or a derivative derived from it (in particular a functional derivative) is preferably suitable. An EWS derivative is especially preferred which has the amino acids 319-656 of the sequence described in sequence ID No. 1, and in particular a fragment which comprises these amino acids.

Accordingly, the invention also relates to the use of EWS or its derivatives for identifying and characterizing substances that influence the activity of NR.

The invention furthermore relates to the use of nucleic acids with at least 70% homology with sequence ID No. 1 or the sequence range 8-2032 or the sequence range 1000-2011 of sequence ID No. 1, for identifying and characterizing substances that influence the activity of nuclear receptors. Preferably, such nucleic acids are cloned in expression cassettes of suitable expression vectors,

especially eukaryotic expression vectors.

The phrase "nucleic acids with at least 70% homology with sequence ID No. 1" is understood in the context of the present invention to mean the entire range between 70% and 100% homology (that is, a complete match with sequence ID No. 1). The selection of a nucleic acid suitable for the intended purpose and within the aforementioned homology range is within the competence of one of conventional skill in the art. The determination of the nucleic acid homology is done in the way familiar to one skilled in this art. For that purpose, various computer programs known to one skilled in the art may be used (such as BLAST; BLAST-2, ALIGN, or Megalign (DNASTAR)).

The method of the invention and the use according to the invention of the aforementioned proteins and nucleic acids are especially suitable for analyzing the hormonal effect of substances on androgen receptors, estrogen receptors (α and β), progesterone receptors (A and B), glucocorticoid receptors, mineral corticoid receptors, thyroid hormone receptors, vitamin D receptors, peroxisome proliferator-activated receptors, retinoic acid receptors, retinoid-X receptors, or orphan receptors. Because of the especially well-characterized action of EWS and EWS derivatives on the androgen receptor, it is used in especially preferred embodiments of the invention.

EWS is furthermore used as a clinical indicator of androgen-conditioned diseases. Relevant androgen-conditioned diseases are for instance prostate cancer, alopecia, acne, or hypogonadism, as well as androgen-resistant syndromes, such as testicular feminization. These are presumed to be due to defects in the co-modulation mechanism between the androgen receptor (AR) and EWS. One plausible diagnostic option for patients with such problems would thus be to measure the
relative rates of AR and EWS. This can be done by using quantitative methods for measuring the relative amount of both molecules in the cellular tissue of the particular patient.

Another aspect of the invention accordingly relates to the use of a nucleic acid

with at least 70% homology with sequence ID No. 1 or the sequence range from 8-2032 or 1000- 2011 of sequence ID No. 1 or of an antibody which is directed against a protein coded by one of these nucleic acids, for diagnosing diseases that are associated with a dysfunction in an NR activity, and preferably androgen receptor activity.

Such an application is preferably employed in the context of a method for determining disturbances in the co- modulation mechanism between an androgen receptor and EWS, in which the cellular concentration or tissue concentrations of the androgen receptor and EWS are measured. Among the various suitable techniques for this for one skilled in the art are particularly radioimmunoassays, ELISAs, immune staining, quantitative RT-PCRs, or Western blots.

Such measurements of the relative rates of AR versus EWS are based on the theory that an androgen-resistance syndrome is based on a disturbance in the equilibrium between AR and EWS prevalence in the target cells. Too much EWS could lead to oversensitivity of the AR system, so that it reacts to molecules that normally have no androgenic effect. Undersensitivity from a lack or malfunction of EWS can lead to androgen resistance at all levels.

It is furthermore possible, with the aid of suitable EWS-cDNA primers, to construct a PCR assay with which in certain patients mutations of the normal DNA sequence can be demonstrated, or transcripts for the Northern blot assay or a DNA for in-situ hybridization assays can be generated.

Demonstrating excessive EWS in a patient would argue for the use of means or provisions to lower the EWS level, for instance by means of antisense nucleic acids against EWS or EWS derivatives, or similar techniques, so as to reduce the EWS titer in that patient under clinical conditions. The same could be achieved by means of molecules that are capable of inhibiting the interaction between AR and EWS.

However, if a patient conversely has an overly low EWS level, then he could be given EWS-cDNA, EWS protein, or EWS- DNA via various mechanisms, in

-10-

order in this way to increase the titer of active EWS. A further aspect of the invention accordingly relates to the use of the aforementioned nucleic acids coding for EWS or EWS derivatives, or EWS proteins and EWS derivatives that are coded by such a nucleic acid, for treating diseases caused by dysfunction in the NR activity.

The invention will be described in further detail below in terms of one example, in conjunction with the drawings.

Shown are:

Fig. 1, a schematic illustration of the gene for the androgen receptor (AR) and the AR2 fragment;

Fig. 2, a schematic illustration of the gene for the Ewing's sarcoma protein (EWS);

Fig. 3, the sequences of EWS exons and EWS proteins;

Fig. 4, the co-activation of the AR signal in SH-SY5Y cells;

Fig. 5, the tissue distribution of the EWS transcript (Fig. 5a) and of the AR transcript (Fig. 5b).

Example 1:

Oligonucleotides used:

Primers for PCR amplification of the library insert:

Act2c5050Eco: gattacgctagcttgggtgg (SEQ ID No. 3)
Act2-4939Xho: gttgaagtgaacttggcgggg (SEQ ID No. 4)

Primers for the amplification of EWS-cDNA to full length:

EWS-8-Sal: gggtcgacggacgttgagagaacgagg (SEQ ID No. 5)

cESW-c2032-Eco: gggaattctgcggggtctctgcatctagtaggg (SEQ ID No. 6)

Sequencing primer:

XII-139a1: gcttgggtggtcatatgg (SEQ ID No. 7)

<u>Vectors used:</u>

pACT2 (gene bank access number U29899) for the library;

pGBT9-derivatives for the probes: pGBT9rev and pGBT(+1)rev (Roder, K.H.; Wolf, S.S.; Schweizer, M., 1996, Analytical Biochemistry 241, 260-62);

pCR2.1 Topo-Vector (made by Invitrogen) for cloning PCR fragments;

CMX-Vector for expression in mammal cells;

pAHluc for the reporter gene assay (contains the MMTV promoter and a luciferase reporter gene; made by A. Cato);

pSG5AR (pSG5 with the human gene for the androgen receptor; gene bank access number AAA51775).

<u>Organisms used:</u>

Yeast strains: Y187 and PJ69-2A

E. coli strain: DH5α

Mammal cells: SH-SY5Y (<u>Deutsche Sammlung von Mikroorganismen und Zellkulturen</u> GmbH (DSMZ): SM ACC209);

PC3 (American type Culture Collection
(ATCC): CRL-1435);

PC3AR: PC3 transfected stably with
pSG5AR (A. Cato, Karlsruhe, Germany).

For identifying novel co-modulators of the androgen receptor, a cDNA library
human cDNA library ("Matchmaker" made by Clontech; No. HY4028AH) from fetal
brain with three different fragments of the androgen receptor (AR) was screened
as a probe, with the aid of the yeast two-hybrid system.

To that end, the vector pSG5AR, which contains the cDNA for the human AR
(gene bank AAA51775) was cleaved with the aid of the endonuclease Psti such
that three different AR-DNA fragments were created. The shortest of these
fragments (AR4) codes for the N-terminus of the receptor (AS 1-56), the middle
one (AR3) codes for the middle part with the activation domains (AS 57-324), and
the longest one (AR2) codes for the C-terminus (AS 325-918) with the DNA and
ligand binding domains (DBD and LBD; see overview in Fig. 1). AR2 was cloned in
the pGBT9(+1)rev vector, once the latter was first linearized with the aid of Psti.

Next, the transformation of the pGBT vector that contained the AR fragment
was done in the yeast strain PJ69- 2A. The positive transformants (Trp+) were
incubated in accordance with the instructions of the manufacturer (Clontech) with
a cDNA library obtained from fetal brain (Humane Multiple-Tissue-cDNA (MTC),
Panel II by Clontech; Catalog No. K1421-1). In accordance with the instructions of
the manufacturer (Clontech), 3×10^6 clones were screened. The positive clones
were selected and tested for their β- galactosidase activity in accordance with the
instructions of the manufacturer (Clontech). The inserts of blue colonies that
originated in the library were amplified by means of PCR and using the primers
Act2c5050Eco and ACT2-4939Xho directly from the yeast cells.

The PCR products were separated by length using gel electrophoresis and
further analyzed by means of cleavage by Mspl. At least one example of each

restriction fragment pattern was sequenced, using XII-139a1 as a sequence primer. The sequences were calibrated with the gene bank or the database Incyte.

One of the multiply identified inserts head a length of 1500 bp and it was possible to identify it (survey in Fig. 2; amino acid sequence in Fig. 3) by sequencing and sequence comparison with the NCBJ database as coding for the C- terminal portion of human EWS (AS 319-656).

Fig. 3 shows the cDNA sequence of human EWS, together with the derived amino acid sequence. Exons 1-17 are shown. The letters printed in bold designate the fragment that can be found in yeast two-hybrid systems and that binds to the AR portions AS 325-AS 919. The sequence ranges that are missing in the slice variants ESW1-b (underlined continuously) and EWS1-c (underlined in dotted lines) are underlined.

By means of PCR and using the primers EWS-8-Sal and cEWS-c2032-Eco, as well as thymus or spleen cDNA made by Clontech, EWS was amplified to full length; the fully coded region of the transcript was isolated from the spleen, and the variant with exon 15 B instead of exon 15 was isolated from the thymus gland. The amplified cDNA was then cloned with EcoRI and Sal I into the expression cassette of the mammal expression vector CMX.

As shown in the bar graph in Fig. 4, EWS after transient transfection in SH-SY5Y cells is capable of bringing about a strong co-activation of the AR signal activity, especially at low androgen concentrations of from 10^{-12} to 10^{-10} mol. To that end, on reaction plates each with six reaction depressions, SH-SY5Y cells with 0.75 µg of a vector that contains a cDNA for the human androgen receptor (pSG5AR), 1.5 µg of the reporter gene construct pAHluc, which contains the MMTV promotor preceding the luciferase gene, and 1 µm of the EWS-CMX vector were co-transfected. The transfection took place using Lipofectin made by Gibeo BRC in accordance with the instructions of the manufacturer. Twenty- four hours after the transfection, the cells were incubated overnight with various quantities of androgen. The cells were lysed with a conventional lysis buffer, and the luciferase activity was measured in the Lumistar luminometer made by BMG Lab

Technologies. The EWS-CMX luciferase activities were compared with the control activities (empty CMX vector) The mixture in each depression was measured in four depressions of a microtiter plate. The control values of the substance without DHT were subtracted. The standard deviation is represented in Fig. 4 by beams.

The tissue distribution of human EWS in normal human tissue can be seen in Fig. 5a, from the autoradiographs shown. For this purpose, an EWS-cDNA fragment that codes for the amino acids 244-656 of EWS was marked with ^{32}P-α-dATP and the Klenow fragment, using the Megaprime^{TM} DNA marking system (Amersham Life Science). The marked fragment was cleaned using a Nick column (Pharmacia) in accordance with the instructions of the manufacturer and hybridized with human blots; Human Northern Blot (MTN) made by Clontech No. 7760-1 and 7759-1) made by Clontech. As can be seen from Fig. 5a, EWS-RNA is expressed predominantly in the testes. Various EWS quantities are also demonstrable in most of the organs examined.

Fig. 5b shows the tissue distribution of human AR in normal human tissue.

From Figs. 5a and 5b, the normal expressions of these two proteins in tissue can be ascertained.

Fig. 4 shows the co-activation of the AR signal in SH- SY5Y cells. In a reaction plate with six reaction depressions, 1 μg of co-activator, 1.5 μg of MMTV-luciferase, and 0.5 μg of pSG5AR plasmid were placed in each depression. The six mixtures were transferred to four depressions each in a microtiter plate and measured there. The error deviation is represented as SD. In all the signals, the values of the corresponding controls without DHT were subtracted.

Fig. 5a shows the tissue distribution of the EWS transcript (Northern Blot MTN made by Clontech). In accordance with the instructions of the manufacturer (Amersham), a random priming of the EWS-cDNA fragment, which codes for the amino acids 244 to 656, and marking with ^{32}P- αdATP and the Klenow fragment were done. The blots were hybridized with the probe, washed, transferred to a film, and developed.

<110> JENAPHARM GmbH & Co. KG

<120> Method for determining the hormonal effect of substances

<130> Pat 3684/11

<140>
<141>

<160> 7

<170> PatentIn Ver. 2.1

<210> 1
<211> 2390
<212> DNA
<213> Homo sapiens

<220> EWS
<221> CDS
<222> (44)..(2011)

<400> 1

```
agagggagac ggacgttgag agaacgagga ggaaggagag aaa atg gcg tcc acg       55
                                                 Met Ala Ser Thr
                                                  1

gat tac agt acc tat agc caa gct gca gcg cag cag ggc tac agt gct      103
Asp Tyr Ser Thr Tyr Ser Gln Ala Ala Ala Gln Gln Gly Tyr Ser Ala
 5                  10                  15                  20

tac acc gcc cag ccc act caa gga tat gca cag acc acc cag gca tat      151
Tyr Thr Ala Gln Pro Thr Gln Gly Tyr Ala Gln Thr Thr Gln Ala Tyr
                 25                   30                  35

ggg caa caa agc tat gga acc tat gga cag ccc act gat gtc agc tat      199
Gly Gln Gln Ser Tyr Gly Thr Tyr Gly Gln Pro Thr Asp Val Ser Tyr
                 40                  45                  50

acc cag gct cag acc act gca acc tat ggg cag acc gcc tat gca act      247
Thr Gln Ala Gln Thr Thr Ala Thr Tyr Gly Gln Thr Ala Tyr Ala Thr
             55                  60                  65

tct tat gga cag cct ccc act ggt tat act act cca act gcc ccc cag      295
Ser Tyr Gly Gln Pro Pro Thr Gly Tyr Thr Thr Pro Thr Ala Pro Gln
         70                  75                  80

gca tac agc cag cct gtc cag ggg tat ggc act ggt gct tat gat acc      343
Ala Tyr Ser Gln Pro Val Gln Gly Tyr Gly Thr Gly Ala Tyr Asp Thr
85                  90                  95                  100
```

```
    acc act gct aca gtc acc acc acc cag gcc tcc tat gca gct cag tct    391
    Thr Thr Ala Thr Val Thr Thr Thr Gln Ala Ser Tyr Ala Ala Gln Ser
                    105             110             115

5
    gca tat ggc act cag cct gct tat cca gcc tat ggg cag cag cca gca    439
    Ala Tyr Gly Thr Gln Pro Ala Tyr Pro Ala Tyr Gly Gln Gln Pro Ala
                    120             125             130

10 gcc act gca cct aca aga ccg cag gat gga aac aag ccc act gag act    487
    Ala Thr Ala Pro Thr Arg Pro Gln Asp Gly Asn Lys Pro Thr Glu Thr
                    135             140             145

    agt caa cct caa tct agc aca ggg ggt tac aac cag ccc agc cta gga    535
15 Ser Gln Pro Gln Ser Ser Thr Gly Gly Tyr Asn Gln Pro Ser Leu Gly
            150             155             160

    tat gga cag agt aac tac agt tat ccc cag gta cct ggg agc tac ccc    583
20 Tyr Gly Gln Ser Asn Tyr Ser Tyr Pro Gln Val Pro Gly Ser Tyr Pro
    165             170             175             180

    atg cag cca gtc act gca cct cca tcc tac cct cct acc agc tat tcc    631
    Met Gln Pro Val Thr Ala Pro Pro Ser Tyr Pro Pro Thr Ser Tyr Ser
25                  185             190             195

    tct aca cag ccg act agt tat gat cag agc agt tac tct cag cag aac    679
    Ser Thr Gln Pro Thr Ser Tyr Asp Gln Ser Ser Tyr Ser Gln Gln Asn
                    200             205             210
30
    acc tat ggg caa ccg agc agc tat gga cag cag agt agc tat ggt caa    727
    Thr Tyr Gly Gln Pro Ser Ser Tyr Gly Gln Gln Ser Ser Tyr Gly Gln
                    215             220             225

35 caa agc agc tat ggg cag cag cct ccc act agt tac cca ccc caa act    775
    Gln Ser Ser Tyr Gly Gln Gln Pro Pro Thr Ser Tyr Pro Pro Gln Thr
            230             235             240

    gga tcc tac agc caa gct cca agt caa tat agc caa cag agc agc agc    823
40 Gly Ser Tyr Ser Gln Ala Pro Ser Gln Tyr Ser Gln Gln Ser Ser Ser
    245             250             255             260

    tac ggg cag cag agt tca ttc cga cag gac cac ccc agt agc atg ggt    871
    Tyr Gly Gln Gln Ser Ser Phe Arg Gln Asp His Pro Ser Ser Met Gly
45                  265             270             275

    gtt tat ggg cag gag tct gga gga ttt tcc gga cca gga gag aac cgg    919
    Val Tyr Gly Gln Glu Ser Gly Gly Phe Ser Gly Pro Gly Glu Asn Arg
                    280             285             290
50
    agc atg agt ggc cct gat aac cgg ggc agg gga aga ggg gga ttt gat    967
    Ser Met Ser Gly Pro Asp Asn Arg Gly Arg Gly Arg Gly Gly Phe Asp
                    295             300             305

55 cgt gga ggc atg agc aga ggt ggg cgg gga gga gga cgc ggt gga atg    1015
    Arg Gly Gly Met Ser Arg Gly Gly Arg Gly Gly Gly Arg Gly Gly Met
            310             315             320
```

17

```
 ggc agc gct gga gag cga ggt ggc ttc aat aag cct ggt gga ccc atg     1063
 Gly Ser Ala Gly Glu Arg Gly Gly Phe Asn Lys Pro Gly Gly Pro Met
 325             330             335             340
5
 gat gaa gga cca gat ctt gat cta ggc cca cct gta gat cca gat gaa     1111
 Asp Glu Gly Pro Asp Leu Asp Leu Gly Pro Pro Val Asp Pro Asp Glu
             345             350             355
10 gac tct gac aac agt gca att tat gta caa gga tta aat gac agt gtg     1159
 Asp Ser Asp Asn Ser Ala Ile Tyr Val Gln Gly Leu Asn Asp Ser Val
             360             365             370
 act cta gat gat ctg gca gac ttc ttt aag cag tgt ggg gtt gtt aag     1207
15 Thr Leu Asp Asp Leu Ala Asp Phe Phe Lys Gln Cys Gly Val Val Lys
             375             380             385
 atg aac aag aga act ggg caa ccc atg atc cac atc tac ctg gac aag     1255
 Met Asn Lys Arg Thr Gly Gln Pro Met Ile His Ile Tyr Leu Asp Lys
20    390             395             400
 gaa aca gga aag ccc aaa ggc gat gcc aca gtg tcc tat gaa gac cca     1303
 Glu Thr Gly Lys Pro Lys Gly Asp Ala Thr Val Ser Tyr Glu Asp Pro
 405             410             415             420
25
 ccc act gcc aag gct gcc gtg gaa tgg ttt gat ggg aaa gat ttt caa     1351
 Pro Thr Ala Lys Ala Ala Val Glu Trp Phe Asp Gly Lys Asp Phe Gln
             425             430             435
30 ggg agc aaa ctt aaa gtc tcc ctt gct cgg aag aag cct cca atg aac     1399
 Gly Ser Lys Leu Lys Val Ser Leu Ala Arg Lys Lys Pro Pro Met Asn
             440             445             450
 agt atg cgg ggt ggt ctg cca ccc cgt gag ggc aga ggc atg cca cca     1447
35 Ser Met Arg Gly Gly Leu Pro Pro Arg Glu Gly Arg Gly Met Pro Pro
             455             460             465
 cca ctc cgt gga ggt cca gga ggc cca gga ggt cct ggg gga ccc atg     1495
40 Pro Leu Arg Gly Gly Pro Gly Gly Pro Gly Gly Pro Gly Gly Pro Met
             470             475             480
 ggt cgc atg gga ggc cgt gga gga gat aga gga ggc ttc cct cca aga     1543
 Gly Arg Met Gly Gly Arg Gly Gly Asp Arg Gly Gly Phe Pro Pro Arg
45 485             490             495             500
 gga ccc cgg ggt tcc cga ggg aac ccc tct gga gga gga aac gtc cag     1591
 Gly Pro Arg Gly Ser Arg Gly Asn Pro Ser Gly Gly Gly Asn Val Gln
             505             510             515
50
 cac cga gct gga gac tgg cag tgt ccc aat ccg ggt tgt gga aac cag     1639
 His Arg Ala Gly Asp Trp Gln Cys Pro Asn Pro Gly Cys Gly Asn Gln
             520             525             530
55 aac ttc gcc tgg aga aca gag tgc aac cag tgt aag gcc cca aag cct     1687
 Asn Phe Ala Trp Arg Thr Glu Cys Asn Gln Cys Lys Ala Pro Lys Pro
             535             540             545
```

/8

```
   gaa ggc ttc ctc ccg cca ccc ttt ccg ccc ccg ggt ggt gat cgt ggc   1735
   Glu Gly Phe Leu Pro Pro Pro Phe Pro Pro Pro Gly Gly Asp Arg Gly
       550             555             560

   aga ggt ggc cct ggt ggc atg cgg gga gga aga ggt ggc ctc atg gat   1783
   Arg Gly Gly Pro Gly Gly Met Arg Gly Gly Arg Gly Gly Leu Met Asp
   565             570             575             580

   cgt ggt ggt ccc ggt gga atg ttc aga ggt ggc cgt ggt gga gac aga   1831
   Arg Gly Gly Pro Gly Gly Met Phe Arg Gly Gly Arg Gly Gly Asp Arg
               585             590             595

   ggt ggc ttc cgt ggt ggc cgg ggc atg gac cga ggt ggc ttt ggt gga   1879
   Gly Gly Phe Arg Gly Gly Arg Gly Met Asp Arg Gly Gly Phe Gly Gly
           600             605             610

   gga aga cga ggt ggc cct ggg ggg ccc cct gga cct ttg atg gaa cag   1927
   Gly Arg Arg Gly Gly Pro Gly Gly Pro Pro Gly Pro Leu Met Glu Gln
           615             620             625

   atg gga gga aga aga gga gga cgt gga gga cct gga aaa atg gat aaa   1975
   Met Gly Gly Arg Arg Gly Gly Arg Gly Gly Pro Gly Lys Met Asp Lys
       630             635             640

   ggc gag cac cgt cag gag cgc aga gat cgg ccc tac tagatgcaga        2021
   Gly Glu His Arg Gln Glu Arg Arg Asp Arg Pro Tyr
   645             650             655

gaccccgcag agctgcattg actaccagat ttatttttta aaccagaaaa tgtttaaat    2081

ttataattcc atatttataa tgttggccac aacattatga ttattccttg tctgtacttt    2141

agtatttttc accatttgtg aagaaacatt aaaacaagtt aaatggtagt gtgcggagtt    2201

ttttttctt ccttctttta aaaatggttg tttaagactt taacaatggg aacccttgt     2261

gagcatgctc agtatcattg tggagaacca agagggcctc ttaactgtaa caatgttcat    2321

ggttgtgatg tttttttttt tttttaaaa taaaattcca aatgtttaat aaaaaaaaa    2381

aaaaaaaaa                                                           2390


<210> 2
<211> 656
<212> PRT
<213> Homo sapiens

<400> 2
Met Ala Ser Thr Asp Tyr Ser Thr Tyr Ser Gln Ala Ala Ala Gln Gln
1               5               10              15

Gly Tyr Ser Ala Tyr Thr Ala Gln Pro Thr Gln Gly Tyr Ala Gln Thr
            20              25              30
```

19

```
   Thr Gln Ala Tyr Gly Gln Gln Ser Tyr Gly Thr Tyr Gly Gln Pro Thr
           35                      40                  45

   Asp Val Ser Tyr Thr Gln Ala Gln Thr Thr Ala Thr Tyr Gly Gln Thr
 5         50                  55                  60

   Ala Tyr Ala Thr Ser Tyr Gly Gln Pro Pro Thr Gly Tyr Thr Thr Pro
   65                      70                  75                  80

10 Thr Ala Pro Gln Ala Tyr Ser Gln Pro Val Gln Gly Tyr Gly Thr Gly
                   85                  90                  95

   Ala Tyr Asp Thr Thr Thr Ala Thr Val Thr Thr Thr Gln Ala Ser Tyr
                100                 105                 110
15
   Ala Ala Gln Ser Ala Tyr Gly Thr Gln Pro Ala Tyr Pro Ala Tyr Gly
           115                 120                 125

   Gln Gln Pro Ala Ala Thr Ala Pro Thr Arg Pro Gln Asp Gly Asn Lys
20     130                 135                 140

   Pro Thr Glu Thr Ser Gln Pro Gln Ser Ser Thr Gly Gly Tyr Asn Gln
   145                 150                 155                 160

25 Pro Ser Leu Gly Tyr Gly Gln Ser Asn Tyr Ser Tyr Pro Gln Val Pro
                   165                 170                 175

   Gly Ser Tyr Pro Met Gln Pro Val Thr Ala Pro Pro Ser Tyr Pro Pro
                180                 185                 190
30
   Thr Ser Tyr Ser Ser Thr Gln Pro Thr Ser Tyr Asp Gln Ser Ser Tyr
                195                 200                 205

   Ser Gln Gln Asn Thr Tyr Gly Gln Pro Ser Ser Tyr Gly Gln Gln Ser
35     210                 215                 220

   Ser Tyr Gly Gln Gln Ser Ser Tyr Gly Gln Gln Pro Pro Thr Ser Tyr
   225                 230                 235                 240

40 Pro Pro Gln Thr Gly Ser Tyr Ser Gln Ala Pro Ser Gln Tyr Ser Gln
                   245                 250                 255

   Gln Ser Ser Ser Tyr Gly Gln Gln Ser Ser Phe Arg Gln Asp His Pro
                260                 265                 270
45
   Ser Ser Met Gly Val Tyr Gly Gln Glu Ser Gly Gly Phe Ser Gly Pro
                275                 280                 285

   Gly Glu Asn Arg Ser Met Ser Gly Pro Asp Asn Arg Gly Arg Gly Arg
50     290                 295                 300

   Gly Gly Phe Asp Arg Gly Gly Met Ser Arg Gly Gly Arg Gly Gly Gly
   305                 310                 315                 320

55 Arg Gly Gly Met Gly Ser Ala Gly Glu Arg Gly Gly Phe Asn Lys Pro
                   325                 330                 335
```

20

```
    Gly Gly Pro Met Asp Glu Gly Pro Asp Leu Asp Leu Gly Pro Pro Val
                340             345             350

    Asp Pro Asp Glu Asp Ser Asp Asn Ser Ala Ile Tyr Val Gln Gly Leu
  5         355             360             365

    Asn Asp Ser Val Thr Leu Asp Asp Leu Ala Asp Phe Phe Lys Gln Cys
            370             375             380

 10 Gly Val Val Lys Met Asn Lys Arg Thr Gly Gln Pro Met Ile His Ile
    385             390             395             400

    Tyr Leu Asp Lys Glu Thr Gly Lys Pro Lys Gly Asp Ala Thr Val Ser
                405             410             415
 15
    Tyr Glu Asp Pro Pro Thr Ala Lys Ala Ala Val Glu Trp Phe Asp Gly
                420             425             430


 20 Lys Asp Phe Gln Gly Ser Lys Leu Lys Val Ser Leu Ala Arg Lys Lys
                435             440             445

    Pro Pro Met Asn Ser Met Arg Gly Gly Leu Pro Pro Arg Glu Gly Arg
        450             455             460
 25
    Gly Met Pro Pro Pro Leu Arg Gly Gly Pro Gly Gly Pro Gly Gly Pro
    465             470             475             480

    Gly Gly Pro Met Gly Arg Met Gly Gly Arg Gly Gly Asp Arg Gly Gly
 30             485             490             495

    Phe Pro Pro Arg Gly Pro Arg Gly Ser Arg Gly Asn Pro Ser Gly Gly
                500             505             510

 35 Gly Asn Val Gln His Arg Ala Gly Asp Trp Gln Cys Pro Asn Pro Gly
                515             520             525

    Cys Gly Asn Gln Asn Phe Ala Trp Arg Thr Glu Cys Asn Gln Cys Lys
        530             535             540
 40
    Ala Pro Lys Pro Glu Gly Phe Leu Pro Pro Pro Phe Pro Pro Pro Gly
   .545             550             555             560

    Gly Asp Arg Gly Arg Gly Gly Pro Gly Gly Met Arg Gly Gly Arg Gly
 45             565             570             575

    Gly Leu Met Asp Arg Gly Gly Pro Gly Gly Met Phe Arg Gly Gly Arg
                580             585             590

 50 Gly Gly Asp Arg Gly Gly Phe Arg Gly Gly Arg Gly Met Asp Arg Gly
                595             600             605

    Gly Phe Gly Gly Gly Arg Arg Gly Gly Pro Gly Gly Pro Pro Gly Pro
            610             615             620
 55
    Leu Met Glu Gln Met Gly Gly Arg Arg Gly Gly Arg Gly Gly Pro Gly
    625             630             635             640
```

Lys Met Asp Lys Gly Glu His Arg Gln Glu Arg Arg Asp Arg Pro Tyr
                    645   650   655

5
  <210> 3
  <211> 20
  <212> DNA
  <213> synthetisch
10
  <220> Primer
  <400> 3
  gattacgcta gcttgggtgg   20

15
  <210> 4
  <211> 21
  <212> DNA
  <213> synthetisch
20
  <220> Primer
  <400> 4
  gttgaagtga acttggcggg g   21

25
  <210> 5
  <211> 27
  <212> DNA
  <213> synthetisch
30
  <220> Primer
  <400> 5
  gggtcgacgg acgttgagag aacgagg   27

35
  <210> 6
  <211> 33
  <212> DNA
  <213> synthetisch
40
  <220> Primer
  <400> 6
  gggaattctg cggggtctct gcatctagta ggg   69

45
  <210> 7
  <211> 18
  <212> DNA
  <213> synthetisch
50
  <220> Primer
  <400> 7
  gcttgggtgg tcatatgg   17

55

Claims

1. A method for determining the hormonal effect of substances, having the following steps:

a) putting EWS protein or a derivative thereof in contact with NR or a derivative thereof and a test substances, and

b) determining the influence of the test substance on the binding of EWS protein or its derivative and NR or its derivative; or

c) determining the influence of the test substance on the ligand-induced activity of the nuclear receptor.

2. The method as defined by claim 1, characterized in that at least one step is performed in the cellular system.

3. The method as defined by claim 2, having the following steps:

a) cells which express EWS protein or a derivative thereof and NR or a derivative thereof are exposed to the substance to be tested;

b) determining the influence of the substance on the interaction between the receptor or its fragment and EWS protein or a derivative by measuring the protein-protein interaction or the protein-protein DNA interaction.

4. The method as defined by claim 2, having the following steps:

a) cells which express EWS protein or a derivative thereof and NR or a derivative thereof and are transfected with a reporter gene construct are exposed to the ligand of the NR and to the substance to be tested;

b) determining the transcription activity of the NR by measuring the reporter

23

gene activity; and

c) comparison with the transcription activity in performing steps a) and b) in the absence of the substance to be tested.

5. The method as defined by claim 3 or 4, characterized in that the cells are eukaryotic cells, preferably prostate cells, nerve cells, glial cells, fibroblasts, blood cells, osteoblasts, osteoclasts, hepatocytes, epithelial cells, or muscle cells.

6. The use of EWS protein or a derivative thereof, which has the function of modulating the activity of at least one NR, for identifying and characterizing substances that influence the activity of NR.

7. The method as defined by one of claims 1-4 or the use as defined by claim 6, characterized in that the EWS derivative includes a derivative, created by amino acid deletion, substitution, insertion, inversion, addition, or exchange, of the polypeptide coded by the nucleic acid sequence in accordance with sequence ID No. 1.

8. The use of a nucleic acid coded for EWS or for an EWS derivative for identifying and characterizing substances which influence the activity of NR.

9. The use as defined by claim 8, characterized in that the nucleic acid has at least 70% homology with sequence ID No. 1 or the sequence range 8-2032 or the sequence range 1000-2001 of the sequence ID No. 1.

10. The use as defined by one of claims 8 or 9, characterized in that the nucleic acid is cloned into the expression cassette of an expression vector.

11. The method as defined by one of claims 1 through 4, or the use as defined by one of claims 6 through 10, characterized in that the NR is an androgen receptor, estrogen receptor, progesterone receptor, glucocorticoid receptor, mineral corticoid receptor, thyroid hormone receptor, vitamin D receptor, peroxisome proliferator- activated receptor, retinoic acid receptor, retinoid-X

receptor, or an orphan receptor, and is preferably an androgen receptor.

12. The use of a nucleic acid with at least 70% homology with sequence ID No. 1 or the sequence range 8-2032 or 1000-2011 of sequence ID No. 1 or of an antibody which is directed against a protein coded by one of these nucleic acids, for diagnosing diseases which are associated with a dysfunction in an NR activity, preferably an activity of the androgen receptor.

13. The use of one of the nucleic acids of claim 12, of a protein which is coded by such a nucleic acid, or of an antisense nucleic acid directed against such a nucleic acid, for the treatment of diseases caused by the dysfunction in the NR activity.

14. A method for determining disturbances in the co- modulation mechanism between an androgen receptor and EWS, wherein the cellular concentrations and/or tissue concentration of the androgen receptor and EWS is measured.

15. The method as defined by claim 14, wherein the measurement of concentration is done by radioimmunoassay, ELISA, immune staining, RT-PCR Western blot, or Northern blot.

25

Abstract

The invention relates to a method for determining the hormonal effect of substances, having the steps of a) putting EWS (Ewing's sarcoma protein) or an EWS derivative thereof in contact with a nuclear receptor (NR) or a derivative thereof and a test substance, and b) determining the influence of the test substance on the binding of EWS protein or its derivative and NR or its derivative, or c) determining the influence of the test substance on the ligand-induced activity of the NR. The invention also relates to a method for determining disturbances in the co-modulation mechanism between an androgen receptor and EWS, which includes measuring the cellular concentrations of androgen receptor and EWS, and to the use of EWS or derivatives for identifying and characterizing substances that affect the activity of NR.

26
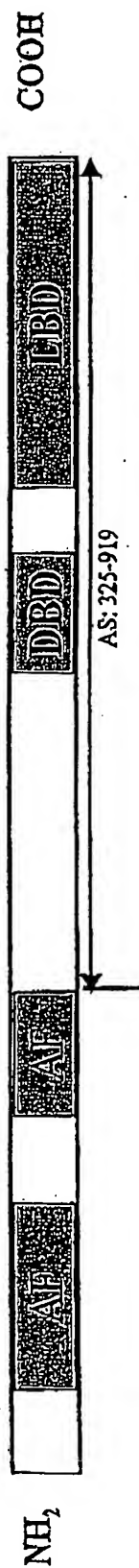
Fig. 1: Schematic representation of the gene for the androgen receptor (AR) and the AR2 fragment

AS 325-919: Androgen receptor fragment 2

AS: Amino acid(s)

AF: Activation domain

DBD: DNA binding domain

LBD: Ligand binding domain
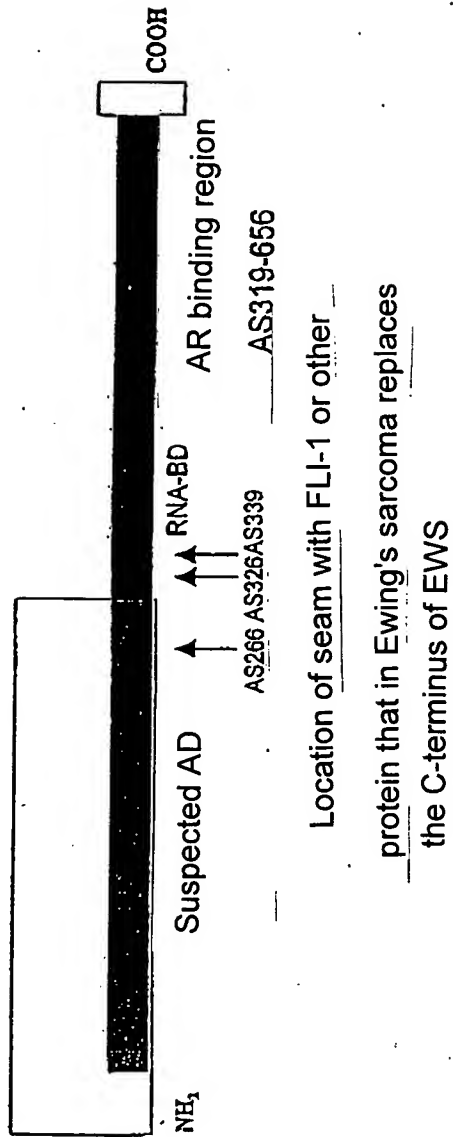
Fig. 2: Schematic representation of the gene for the EWS protein (EWS)

Blue: RNA binding domain

Dark red: Androgen receptor binding region (AS 319- 656)

AS: Amino acid(s)

AD: Activation domain

BD: Binding domain

```
  1  agagggagac ggacgttgag agaacgagga ggaaggagag aaaATGGCGT CCACGGATTA
                                                       M   A   S   T   D
                                                                    >>>>
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 1>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

 61  CAGTACCTAT AGCCAAGCTG CAGCGCAGCA GGGCTACAGT GCTTACACCG CCCAGCCCAC
      Y   S   T   Y   S   Q   A   A   A   Q   Q   G   Y   S   A   Y   T   A   Q   P
     >>>>>>>>>>>>exon 2>>>>>>>>>>>>>>>
                                        >>>>>>>>>>>exon 3>>>>>>>>>>

121  TCAAGGATAT GCACAGACCA CCCAGGCATA TGGGCAACAA AGCTATGGAA CCTATGGACA
      T   Q   G   Y   A   Q   T   T   Q   A   Y   G   Q   Q   S   Y   G   T   Y   G
                          >>>>>>>>>>>>>>>>exon 4>>>>>>>>>>>>>>>>>
     >>>>>>>>>>>exon 3>>>>>>>>>>>

181  GCCCACTGAT GTCAGCTATA CCCAGGCTCA GACCACTGCA ACCTATGGGC AGACCGCCTA
      Q   P   T   D   V   S   Y   T   Q   A   Q   T   T   A   T   Y   G   Q   T   A
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 4>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                       q

241  TGCAACTTCT TATGGACAGC CTCCCACTGG TTATACTACT CCAACTGCCC CCCAGGCATA
      Y   A   T   S   Y   G   Q   P   P   T   G   Y   T   T   P   T   A   P   Q   A
                          >>>>>>>>>>>>>>exon 5>>>>>>>>>>>>>>>>
     >>>>>>>>>>>exon 4>>>>>>>>>>>>>

301  CAGCCAGCCT GTCCAGGGGT ATGGCACTGG TGCTTATGAT ACCACCACTG CTACAGTCAC
      Y   S   Q   P   V   Q   G   Y   G   T   G   A   Y   D   T   T   T   A   T   V
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 5>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

361  CACCACCCAG GCCTCCTATG CAGCTCAGTC TGCATATGGC ACTCAGCCTG CTTATCCAGC
      T   T   T   Q   A   S   Y   A   A   Q   S   A   Y   G   T   Q   P   A   Y   P
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 5>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

421  CTATGGGCAG CAGCCAGCAG CCACTGCACC TACAAGACCG CAGGATGGAA ACAAGCCCAC
      A   Y   G   Q   Q   P   A   A   T   A   P   T   R   P   Q   D   G   N   K   P
                               >>>>>>>>>>exon 6>>>>>>>>>>
     >>>>>>>>>>>>>>>exon 5>>>>>>>>>>>>>>>>

481  TGAGACTAGT CAACCTCAAT CTAGCACAGG GGGTTACAAC CAGCCCAGCC TAGGATATGG
      T   E   T   S   Q   P   Q   S   S   T   G   G   Y   N   Q   P   S   L   G   Y
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 6>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

541  ACAGAGTAAC TACAGTTATC CCCAGGTACC TGGGAGCTAC CCCATGCAGC CAGTCACTGC
      G   Q   S   N   Y   S   Y   P   Q   V   P   G   S   Y   P   M   Q   P   V   T
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 6>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
601  ACCTCCATCC TACCCTCCTA CCAGCTATTC CTCTACACAG CCGACTAGTT ATGATCAGAG
      A   P   P   S   Y   P   P   T   S   Y   S   S   T   Q   P   T   S   Y   D   Q
                              >>>>>>>>>>>>>>>>>>exon 7>>>>>>>>>>>>>>>>>>>
     >>>>>>>>>exon 6>>>>>>>>>>
```

```
661  CAGTTACTCT CAGCAGAACA CCTATGGGCA ACCGAGCAGC TATGGACAGC AGAGTAGCTA
      S   S   Y   S   Q   Q   N   T   Y   G   Q   P   S   S   Y   G   Q   Q   S   S
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 7>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
721  TGGTCAACAA AGCAGCTATG GGCAGCAGCC TCCCACTAGT TACCCACCCC AAACTGGATC
      Y   G   Q   Q   S   S   Y   G   Q   Q   P   P   T   S   Y   P   P   Q   T   G
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 7>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
781  CTACAGCCAA GCTCCAAGTC AATATAGCCA ACAGAGCAGC AGCTACGGGC AGCAGAGTTC
      S   Y   S   Q   A   P   S   Q   Y   S   Q   Q   S   S   S   Y   G   Q   Q   S
                                                                            >>>>
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 7>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
841  ATTCCGACAG GACCACCCCA GTAGCATGGG TGTTTATGGG CAGGAGTCTG GAGGATTTTC
      S   F   R   Q   D   H   P   S   S   M   G   V   Y   G   Q   E   S   G   G   F
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 8>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
901  CGGACCAGGA GAGAACCGGA GCATGAGTGG CCCTGATAAC CGGGGCAGGG GAAGAGGGGG
      S   G   P   G   E   N   R   S   M   S   G   P   D   N   R   G   R   G   R   G
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 8>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
961  ATTTGATCGT GGAGGCATGA GCAGAGGTGG GCGGGGAGGA GGACGCGGTG GAATGGGCAG
      G   F   D   R   G   G   M   S   R   G   G   R   G   G   G   R   G   G   M   G
                                                                            >>>
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 8>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
1021 CGCTGGAGAG CGAGGTGGCT TCAATAAGCC TGGTGGACCC ATGGATGAAG GACCAGATCT
      S   A   G   E   R   G   G   F   N   K   P   G   G   P   M   D   E   G   P   D
                                          >>>>>>>>>>>exon 10>>>>>>>>>>
     >>>>>>>>>>>>>>>exon 9>>>>>>>>>>>>>>>>>
```

```
1081 TGATCTAGGC CCACCTGTAG ATCCAGATGA AGACTCTGAC AACAGTGCAA TTTATGTACA
      L   D   L   G   P   P   V   D   P   D   E   D   S   D   N   S   A   I   Y   V
                >>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 11>>>>>>>>>>>>>>>>>>>>>>>>>>
     >>>>>>>>
```

```
1141 AGGATTAAAT GACAGTGTGA CTCTAGATGA TCTGGCAGAC TTCTTTAAGC AGTGTGGGGT
      Q   G   L   N   D   S   V   T   L   D   D   L   A   D   F   F   K   Q   C   G
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 11>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

```
1201  TGTTAAGATG AACAAGAGAA CTGGGCAACC CATGATCCAC ATCTACCTGG ACAAGGAAAC
       V  V  K  M   N  K  R   T  G  Q   P  M  I  H   I  Y  L   D  K  E
          >>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 12>>>>>>>>>>>>>>>>>>>>>>>>>>>>
          >>>>>>>


1261  AGGAAAGCCC AAAGGCGATG CCACAGTGTC CTATGAAGAC CCACCCACTG CCAAGGCTGC
       T  G  K  P   K  G  D   A  T  V   S  Y  E  D   P  P  T   A  K  A
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 12>>>>>>>>>>>>>>>>>>>>>>>>>>>>


1321  CGTGGAATGG TTTGATGGGA AAGATTTTCA AGGGAGCAAA CTTAAAGTCT CCCTTGCTCG
       A  V  E  W   F  D  G   K  D  F   Q  G  S   K  L  K  V   S  L  A
                    >>>>>>>>>>>>>>>>>>>>>>>exon 13>>>>>>>>>>>>>>>>>>>>>>>>
       >>>>>exon 12>>>>>>


1381  GAAGAAGCCT CCAATGAACA GTATGCGGGG TGGTCTGCCA CCCCGTGAGG GCAGAGGCAT
       R  K  K  P   P  M  N   S  M  R   G  G  L   P  P  R  E   G  R  G
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 13>>>>>>>>>>>>>>>>>>>>>>>>>>>>>


1441  GCCACCACCA CTCCGTGGAG GTCCAGGAGG CCCAGGAGGT CCTGGGGGAC CCATGGGTCG
       M  P  P  P   L  R  G   G  P  G   G  P  G  G   P  G  G   P  M  G
                     >>>>>>>>>>>>>>>>>>>exon 14>>>>>>>>>>>>>>>>>>>
       >>>>>>exon 13>>>>>>>>>


1501  CATGGGAGGC CGTGGAGGAG ATAGAGGAGG CTTCCCTCCA AGAGGACCCC GGGGTTCCCG
       R  M  G  G   R  G  G   D  R  G   G  F  P  P   R  G  P   R  G  S
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 14>>>>>>>>>>>>>>>>>>>>>>>>>>>>>


1561  AGGGAACCCC TCTGGAGGAG GAAACGTCCA GCACCGAGCT GGAGACTGGC AGTGTCCCAA
       R  G  N  P   S  G  G   G  N  V   Q  H  R   A  G  D  W   Q  C  P
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 14>>>>>>>>>>>>>>>>>>>>>>>>>>>>>


1621  TCCGGGTTGT GGAAACCAGA ACTTCGCCTG GAGAACAGAG TGCAACCAGT GTAAGGCCCC
       N  P  G  C   G  N  Q   N  F  A   W  R  T   E  C  N  Q   C  K  A
          >>>>>>>>>>>>>>>>>exon 15b>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
          >>>>>>>>>>>>>>>>>>>>>>>>>>>exon 15>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
          >>>


1681  AAAGCCTGAA GGCTTCCTCC CGCCACCCTT TCCGCCCCCG GGTGGTGATC GTGGCAGAGG
       P  K  P  E   G  F  L   P  P  P   F  P  P   P  G  G  D   R  G  R
                                                     >>>>>>>exon 16>>>>>>
       >>>>>>>>>>>>>>>>>>>exon 15>>>>>>>>>>>>>>>>>>>>>>
```

1741 TGGCCCTGGT GGCATGCGGG GAGGAAGAGG TGGCCTCATG GATCGTGGTG GTCCCGGTGG
      G   G   P   G   G   M   R   G   G   R   G   G   L   M   D   R   G   G   P   G
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 16>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

1801 AATGTTCAGA GGTGGCCGTG GTGGAGACAG AGGTGGCTTC CGTGGTGGCC GGGGCATGGA
      G   M   F   R   G   G   R   G   G   D   R   G   G   F   R   G   G   R   G   M
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 16>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

1861 CCGAGGTGGC TTTGGTGGAG GAAGACGAGG TGGCCCTGGG GGGCCCCCTG GACCTTTGAT
      D   R   G   G   F   G   G   G   R   R   G   G   P   G   G   P   F   G   P   L
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 16>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

1921 GGAACAGATG GGAGGAAGAA GAGGAGGACG TGGAGGACCT GGAAAAATGG ATAAAGGCGA
      M   E   Q   M   G   G   R   R   G   G   R   G   G   P   G   K   M   D   K   G
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 16>>>>>>>>>>>>>>>>>>>>>>>>>>>   >>>>>>

1981 GCACCGTCAG GAGCGCAGAG ATCGGCCCTA ctagatgcag agaccccgca gagctgcatt
      E   H   R   Q   E   R   R   D   R   P   Y   stop
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 17>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

2041 gactaccaga tttattttt aaaccagaaa atgttttaaa tttataattc catatttata
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 17>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

2101 atgttggcca caacattatg attattcctt gtctgtactt tagtatttt caccatttgt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 17>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

2161 gaagaaacat taaaacaagt taaatggtag tgtgcggagt tttttttct tccttctttt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 17>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

2221 aaaaatggtt gtttaagact ttaacaatgg gaacccttg tgagcatgct cagtatcatt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 17>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

2281 gtggagaacc aagagggcct cttaactgta acaatgttca tggttgtgat gttttttttt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>exon 17>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

2341 ttttttaaa ataaaattcc aaatgtttaa taaaaaaaaa aaaaaaaaa
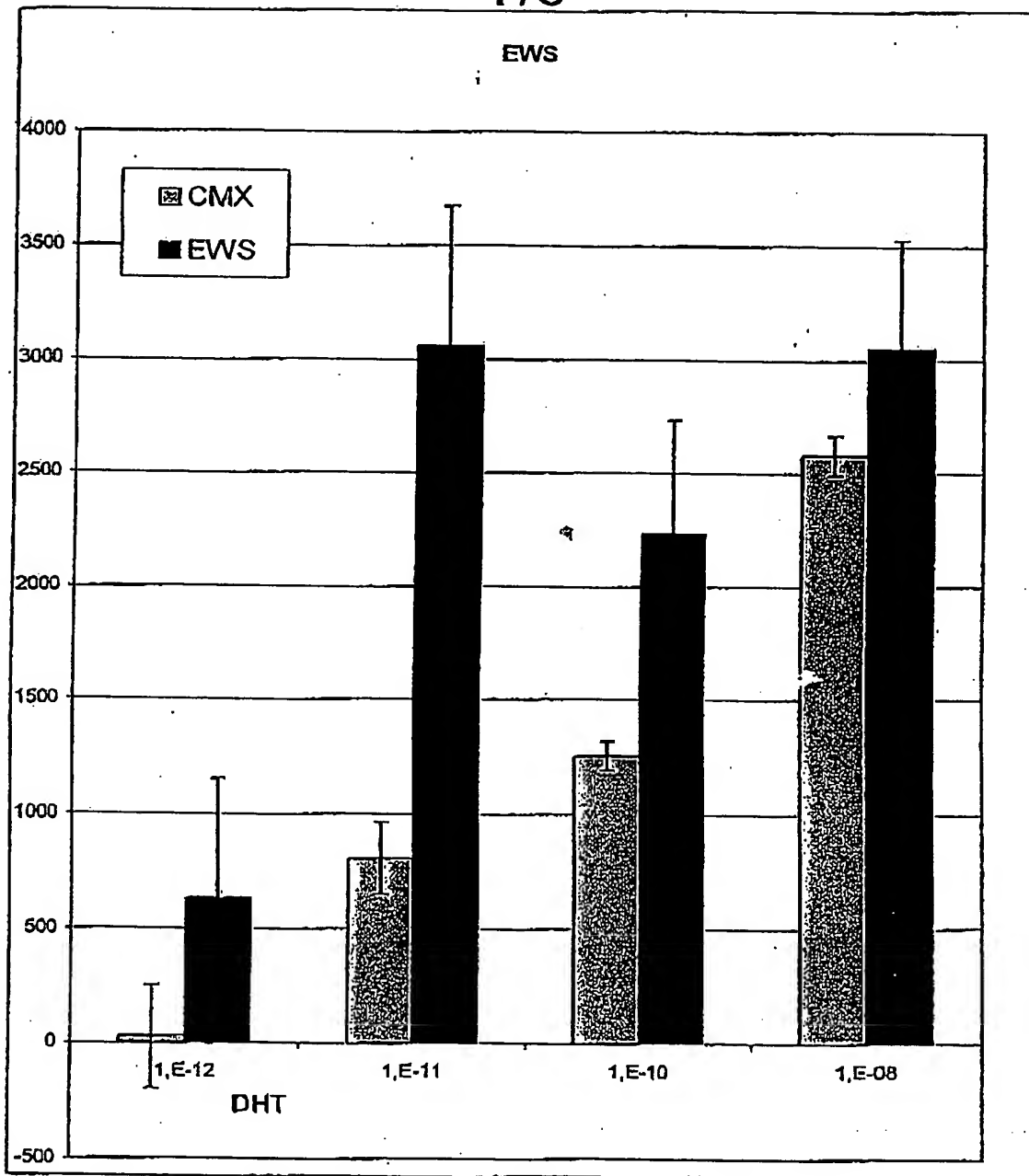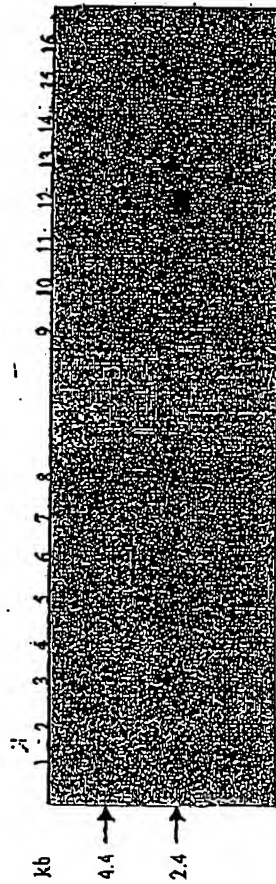   >>>>>>>>>>>>>exon 17>>>>>>>>>>>>>>>

Fig. 4: Co-activation of the AR signal in SH-SY5Y cells

EWS

Fig. 5a

1. Heart	9. Spleen
2. Brain	10. Thymus gland
3. Placenta	11. Prostate
4. Lung	12. Testes
5. Liver	13. Ovary
6. Skeletal muscle	14. Small intestine
7. Kidney	15. Large intestine
8. Pancreas	16. Peripheral leucocytes

Fig. 5b